Theses of Doctoral (Ph.D.) Dissertation

Population dynamics studies of wild boar (*Sus scrofa*) populations in the Carpathian Basin, especially their origin, genetic diversity and geographic isolation

Bendegúz Mihalik
Ph.D. candidate

Dissertation supervisor: Dr. Szilvia Kusza

Dissertation co-supervisor: Dr. Viktor Stéger

UNIVERSITY OF DEBRECEN
Doctoral School of Animal Science

Debrecen, 2022

1. Introduction

Wild boar (*Sus scrofa*) is one of the most widespread big game species in Europe and worldwide. This large, omnivorous and generalist species is the only R strategist big game in Hungary, which means it has many offspring with a short maturation period. Its population has grown steadily since the 1960s until the period of the African swine fever virus (ASF) epidemic. It has doubled its population in Hungary and across Europe on an average of 10-20 years; before the outbreak of ASF in 2018, its population in Hungary was estimated at almost 120,000 individuals. Its size and mobility are both large, with a movement range of 100-15000 hectares. Due to these traits, wild boar have a serious impact on their habitat, both ecologically and economically. The disturbance they bring about has been shown to both reduce and increase biodiversity in Hungary, as well as to cause significant damage in conservation, wildlife as well as agriculture, such as the consumption of eggs from ground-nesting birds, or foraging freshly sown maize. In addition, wild boars cause a large number of accidents: 627 wild boars were hit on Hungarian roads in 2019, which not only caused material damage but also endangered human lives. The excellent adaptability of wild boar is exemplified by the fact that the International Union for Conservation of Nature (IUCN) classified it as one of the 100 most invasive species in the world.

According to the results of previous studies of other species, the Carpathian Basin's populations are divided into two haplotypes, therefore a survey of the Hungarian wild boar population with a representative sample set may prove interesting. The number of studies in Hungary based on modern genetic methods is low, and those few usually used only a few dozen individuals from Hungary, from which the genetic background, diversity, and geographical distribution of the wild boar populations could not be established. In my dissertation I would like to fill this gap, by studying and analyzing the individuals from the Carpathian Basin with modern genetic methods and software to get answers to these questions, which may lead to an even more science-based management of the species.

Research of the genetic basis is an important milestone in the accurate knowledge of the species, as diversity influences adaptation, a large population of individuals with diverse genetic background has a much higher chance of survival in the course of climatic, human, or other events that cause negative effects. In addition, during the management or breeding of a group of animals with a broad genetic background, people are more likely to find favorable variants that they want to spread in that group. Several methods are

known in domestic species (for example cross-breeding) in which certain qualitative or quantitative traits are improved by involving a foreign breed. As wild boars are more resistant than domestic pigs due to their extensive lifestyle, such traits of a group with a good genetic background can be used in the breeding program of a domestic breed, as the wild boar also functions as a „living gene bank". Furthermore, although the wild boar is essentially a game species, it is kept and bred in farms, therefore their disease resistance, meat quality, or even reproductive success, litter size, nutrient utilization and many other properties that are beneficial to domestic animals can be recorded and improved through genetic-based, selected breeding. Finally, under appropriate control and indicating the origin of the product, the breeding of hybrids may be advantageous in some cases, for example, to make meat products with a higher fat content (sausages).

Hybridization between wild boar and domestic pigs is extremely easy, because they belong to a single species. This is further reinforced by the practice of extensive animal husbandry, which was common until the 1970-s, which made it easy for individuals from the two groups to meet and give rise to hybrid offspring in nature. Although hybridization can cause an increase in biodiversity and thus be considered desirable in some cases, it should be avoided in the case of a cross between a game species and a domestic animal. First of all, a domestic breed is selected and has evolved to its present form in an unnatural way, without selection pressure from environmental factors, thus its survival in the wild is weaker than that of its wild counterparts. A good example is the thick and strong skin and fur of the wild boar, as opposed to the almost bare skin of most domestic breeds. Domestic pigs have such a thick fat layer that in many cases it hinders their movement and puts a heavy strain on their hearts. The stress tolerance of some domestic breeds is also extremely low, which is also a serious disadvantage for them in the natural environment, where they are unfamiliar. In addition, the appearance of domestic genes in wild boar populations leads to a decline in the wild gene pool, some alleles are expected to disappear, which causes a decrease in the wild boar gene pool. This is a typical problem in Hungary as well, where gene stocks have been mixed for decades due to the extensive keeping of domestic pigs. In some countries (USA, Australia) the semi-domesticated „feral pig" is present in millions, causing huge damages to agriculture. Finally, it is not desirable for hunters to hunt for specimens that look like domestic pigs, which also causes economic losses to the local game managers.

Last, but not least, accurate knowledge of the genetic background may pave the way for long-term future plans such as preserving the full diversity of a species in gene banks in the event of a disaster or breeding ASF and other pathogen-tolerant individuals.

2. Aim of the study

The main research goals of our work were the following:

- Basic research of the population genetic profile of wild boar populations in the Carpathian Basin, especially in Hungary, including calculation of allele numbers, rates of genetic diversity, heterozygosity and Hardy-Weinberg Equilibrium values.
- Determining the size and geographic distribution of populations/subpopulations, establishing the factors that separate them from each other, and measuring the level of genetic variance between them.
- Testing a marker set that was developed for individual identification on a large number of samples at a regional level.
- Assessing the past effect on the wild boar population of Hungary through population dynamics analyses.
- Testing a wild boar species-specific marker set that was originally developed for origin testing on natural groups. Comparing the parameters (reliability, accuracy) of our markers with previous literature data. If the previously mentioned parameters meet the requirements of the discipline, determining the level of hybridization in the Carpathian Basin's wild boars.

3. Material and methods

Sampling

Wild boar samples

For our study, we collected wild boar muscle tissue (n=423) and fur (n=63) samples from all across Hungary (n=470) and from the neighboring countries (Romania n=5, Croatia n=4, Slovakia n=4). As wild boar can be hunted legally in Hungary all year round, we asked professional and sport hunters for sample collection after shooting a specimen, as well as hair and isolated DNA from researchers at other institutions. In addition, I visited game processing companies (Öreglaki Vadfeldolgozó Kft., Villányi-Vad Kft., Sárrét-Vad Kft.) where I was able to collect a large set of samples covering a significant part of the country all at the same time. Individually marked samples were stored in 2 ml Eppendorf tubes containing 1 ml pure ethanol (muscle tissues) or in nylon bags (hair). The exact Global Positioning System (GPS) coordinates of the hunted animal, or if it wasn't possible, the code of the responsible wildlife management unit was assigned to the ID on paper. The data set was digitalized into a Microsoft Excel file and visualised on a Google Map as soon as it was possible. The collected 486 samples were stored at -20°C until processing.

Domestic pig samples

The samples used for the research (pietrain, hampshire, large white, landrace, duroc, duroc x mangalitza, blond mangalitza, swallow-bellied mangalitza (all n=12)) were collected before the start of our studies, thus they were available to us (Szemethy et al., 2020). Fur samples were taken from domestic pigs that were involved in breeding programs for several generations. The samples were stored and handled as previously described.

Recording background data of collected samples

The geographical locations of the individuals were put into a Microsoft Excel sheet. The accurate GPS coordinate was not available for the majority of the individuals, thus in the other cases we used the map of 1445 game management units in Hungary and their centers, edited by the Hungarian Game Management Database, and I recorded the coordinates of the unit's centers as the exact place of the individuals.

Genomic DNA isolation

As the first step in processing, genomic DNA was isolated from the collected samples. For muscle tissue samples the Geneaid Genomic Tissue Kit (Geneaid, Taiwan) and for hair the QIAamp DNA Investigator Kit (QIAGEN, Germany) was used, following the manufacturer's protocols. The quantity and quality of the isolated DNA were checked on a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, USA). For the polymerase chain reactions (PCR) the DNA was uniformly diluted to 15 ng/µl by the addition of distilled water and stored at -20 °C until used.

Amplification of tested regions

Amplification of STR regions for population genetic studies

Population genetic studies were performed according to Lin et al. (2014) with 13 fluorescently labeled tetranucleotide markers, which were optimized for local conditions in the GBI (Genetics and Biotechnology Institute) of the MATE (Hungarian University of Agriculture and Life Sciences) with the help of the staff of the Applied Wild and Farm Animal Genomics Group.

The multiplex marker set was optimized to 20 µl, the required amount of DNA was 45 ng, therefore 3 µl of the homogenized sample was added to each reaction mixture, and supplemented with 10 µl of Qiagen Multiplex PCR Mix, 0.28-0.65 µl of each primer mix, and 1.23 µl distilled water to fill the volume to 20 µl. The mixture was then placed in a LifeECO (Hangzhou Bioer Technology, Hangzhou, China) PCR machine for DNA amplification. PCR conditions were as follows: initial denaturation period at 95°C for 15 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 61°C for 30 s, extension at 72°C for 60 s and a final extension period at 72°C for 90 min.

Amplification of InDel regions for hybridization studies

For the hybridization studies an InDel marker set containing 3 markers, originally developed by the Applied Wild and Farm Animal Genomics Group of GBI MATE (Szemethy et al., 2020) was used, therefore no optimization was required in this case. Amplification was carried out in a 25-µl reaction volume containing 3 µl DNA (15 ng/µl), 12.5 µl Qiagen Multiplex PCR Mix, 0.7-2.05 µl of each primer mix, and 5.35 µl distilled water. The mixture was then placed in a LifeECO (Hangzhou Bioer Technology, Hangzhou, China) PCR machine for DNA amplification. PCR conditions were as follows: initial denaturation period at 95°C for 15 min, followed by 40 cycles of

denaturation at 94°C for 40 s, annealing at 60°C for 40 s, extension at 72°C for 30 s and a final extension period at 72°C for 5 min.

Genotyping of samples

After amplification, the products were checked by 1.5% agarose gel electrophoresis at 100A for 30 minutes. Based on the results, samples were diluted to 1/10 to 1/100 of their original concentrations for fragment analysis. 2 µl of the products were mixed with 10 µl HIDI formamide (Thermo Fisher Scientific, Massachusetts, USA) and 0.3 µl LIZ500 Size Standard (Thermo Fisher Scientific, Massachusetts, USA). Genotyping was performed by the laboratory of BIOMI Kft. on an ABI 3100 genetic analyzer (Applied Biosystems, Massachusetts, USA).

Statistical evaluation

PeacScanner v1.0 (Applied Biosystems, Foster City, CA, USA) was used to analyze the raw data received from BIOMI Kft. The results are written into a Microsoft Excel spreadsheet.

Determining the population genetic profile of wild boar populations in the Carpathian Basin

GenAlEx v.6.5. software (Peakall & Smouse, 2012) was used for genetic testing of the sample set. Using the program, the allele numbers, effective allele numbers, expected and observed heterozygosity values and the deviation from HWE (Hardy-Weinberg Equilibrium) were determined.

Determining the size and distribution of populations/subpopulations, examining the factors that separate them, and the degree of genetic difference between them

The grouping of individuals and the calculation of the degree of genetic difference between them (F_{st}) was made using Geneland v.3.4.2. (Guillot et al., 2012) software. The genetic profile of individuals and the coordinates of the samples were entered as input files. Based on literature data, the running parameters were set to:

- Number of possible populations: 1-10
- Number of repeats: 100000
- Thinning: 100
- Allele frequency model: non-correlated

- spatial model: true
- null-allele model: false
- multiple single runs: no

The genetic position of subpopulations was visualized by principal component analysis using the Past v.2 software (Hammer et al., 2001). The visual representation of the results obtained was placed on a map from Google maps with ArcGis program. To find the possible separating factors (barriers), we first tried to check the map visually, then followed up with Barrier v.2.2. (Manni et al., 2004) software. Finally, the groups were studied separately in GenAlEx v.6.5. software, as already mentioned in the previous subsection.

Testing of a marker setdeveloped for individual identification

According to data in the literature, the STR set we use is also suitable for individual-level identification. We also tested this statement on our sample set. The assay was performed in Colony v2.0.6.6 (Jones & Wang, 2010) software. The results were visualized by the genetic profiles of each individual written by a single number, and all digits were written into a column. The entire table was sorted by digits, and each digit was given a different colored background for better transparency. Based on the allele numbers we found, the marker set can distinguish the number of individuals shown in Figure 1.

$$\left(n_1 + \frac{(n_1 - 1) \times n_1}{2}\right) \times \left(n_2 + \frac{(n_2 - 1) \times n_2}{2}\right) \times \cdots \times (n_N + \frac{(n_N - 1) \times n_N}{2})$$

Figure 1: **The formula used for calcluating the identifiable unique genetic profiles** (Fegyverneki, 2011)

Legend: n = number of alleles found on a given marker, N = number of markers used

Population dynamics studies of wild boar populations in the Carpathian Basin

Investigation of the bottleneck effect and previous genetic shrinkage of the groups

To test the bottleneck effect the Nottleneck v.1.2.02. software (Cornuet & Luikart, 1996) was used, with which the results of 13 microsatellite markers on 486 wild boars were run. The settings were as follows:

- mutation model:
    - Infinite allele model (IAM): no
    - Stepwise mutation model (SMM): yes
    - Two-phase mutation model (TPM): yes
- variance of TPM: 30%
- proportion of SMM in TPM: 70%
- number of iterations: 1000
- statistical tests:
    - sign test: yes
    - standardized difference test: yes
    - Wilcoxon rank sum test: yes
    - mode-shift: yes

Kinship analyses of the tested wild boar sample set

Kinship analyses were examined using the Colony v2.0.6.6 (Jones & Wang, 2010) software. According to the study of the software's programmers, depending on the relatedness and the strength of the family bond it can determine the relationships between individuals with an accuracy of over 90% even in the case of 4-10 microsatellite markers, therefore it can be used reliably for our studies.

The running parameters of the program were as follows:

- analysis type: empirical data analysis
- Mating system:
    - Male polygamy: yes
    - Female polygamy: yes
    - Inbreeding: yes
    - Clones: no
- Species:
    - sexuality: dioecious
    - ploidity: diploid

- Length of run: long
- Analysis method: full-likelihood
- Likelihood precision: high
- Update allele frequency: yes
- sibship scaling: yes
- number of runs: 1
- random number seed: 1234
- Sibship prior:
    - weak prior
    - paternal sibship size: 1
    - maternal sibship size: 1
- characteristics of markers:
    - marker type: codominant
    - allele dropout rate: 0
    - margin of error: 0.0001
    - allele frequency: unknown
- characteristics of parents:
    - probability of father in the sample set: 0.5
    - probability of mother in the sample set: 0.5
    - number of known paternal offspring: 0
    - number of known maternal offspring: 0
    - excluded paternity: 0
    - excluded maternity: 0
    - excluded paternal brotherhood: 0
    - excluded maternal brotherhood: 0

When evaluating the results, only relatives with a probability above 90% were accepted based on literature data and back-testing. In the last step, we measured the geographical distance between the pairs in Google Maps. Since the exact location of most individuals is only known at the game management unit level, the distance was rounded to 10km.

Testing of a set of markers developed for tracing of wild boar meat products on natural groups, hybridization studies

Investigation of the accuracy of the developed marker set by bioinformatics

The marker set that was originally developed for meat product tracing was subjected to a dual purpose pre-test: its accuracy was determined by bioinformatics and compared with the accuracy of other, most commonly used hybridization markers (MC1R, NR6A1). These are based on SNP polymorphisms in genes that determine fur color; for example, Lorenzini et al. (2020) used these methods in their study. The comparison was performed as follows: 11 domestic pig and 12 wild boar full genome files were downloaded from the NCBI (National Center for Biotechnology Information) database, and the primers used were blasted onto them by IGV 2.3.97 (Robinson et al., 2011) software. We next checked whether the given primer pair produced the allele variation corresponding to the individual (light blue for wild boars, gray for domestic pigs). If an allele with a color corresponding to the variety was indicated by the program, a result of 100% was recorded, 50% for a heterozygous mark (dark blue) and 0% for a mismatch. Finally, the results from the 23 genomes were averaged to obtain the accuracy of the markers.

Examination of the hybridization level of the Hungarian wild boar population

To calculate the hybridization level of individuals, Structure 2.3.3 (Pritchard et al., 2000) was used. The study was performed with the 3 InDel markers only, and with the combined InDel and STR marker set. The reason why the InDel marker set is needed is that the STR set was developed for individual identification, and it cannot reliably separate mangalica from wild boars. Since wild boars were also used to create the mangalica breed, the two groups are genetically closer to each other than the mangalica to the other studied breeds (Szemethy et al., 2020).

The software was used with the following settings:

- The number of groups (K) was set to 2 in both cases, as in this case the aim was to separate wild boars from domestic pigs.
- Parameters:
  - length of burning period: 750000
  - Number of MCMC (Markov Chain Monte Carlo) repeats after burning: 250000
- Number of iterations: 5

Based on literature data, we considered hybrids to be those individuals in which the probability of belonging to both groups reached 25%; other individuals were classified as either a pure wild boar or a pure domestic pig. The efficacy of the markers was also tested with simulated F1-F2 wild boar-domestic pig hybrids, and the results were compared by statistical methods. To create artificial hybrids, we used Hybridlab v1.0 (Nielsen et al., 2006) software, which can simulate hybrid genotypes from the genetic profiles of existing individuals. We used previously tested wild boars of the purest genotype (n=12) and created F1 hybrids with 12 pietrains and 4 individuals from each of the 3 mangalica breeds. In the next step, F1 hybrids were crossed back with wild boars and with individuals from the other breed, producing 75% and 25% wild boar offsprings. Finally, individuals in the F1 groups were also crossed.

4. Results

Determining the population genetic profile of wild boar populations in the Carpathian Basin

As the first step of the studies, following the general division of population genetic surveys, we determined the basic genetic characteristics of the individuals and the sample set. This data set contains significant information about the diversity of individuals and the potential threats to them, such as adaptation difficulties or the possibility of allele loss. Population values were determined using the GenAlEx software. The number of alleles per marker ranged from 4 (PigSTR14A) to 14 (PigSTR11B and PigSTR15A), with a mean value of 7.62. According to an earlier study by Costa et al. (2012), using 14 markers that differ from ours, found 3-14 alleles with an average of 6.21; Vernesi et al. (2003) using a third set of 9 markers found 6-12 markers with an average of 8.8. Compared to other studies in other European countries, the genetic diversity of the Hungarian population is in the middle range. Based on these results, we can determine that the marker set can be used with good efficiency for the genetic testing of Hungarian wild boars. It can also be seen that the diversity of the wild boar population in Hungary is moderate compared to that of other European groups, but this may be due to the use of different marker sets. The expected heterozygosity value was higher than that observed in all cases except for the markers PigSTR4B and PigSTR17A, indicating inbreeding. Significant deviations from HWE were obtained for 9 markers, the significance level was $p<0.05$ in the case of PigSTR14B, and in the other 8 cases (PigSTR7B, PigSTR4C, PigSTR11B, PigSTR1B, PigSTR15A, PigSTR5C, PigSTR13E, PigSTR1A) it was $p<0.001$. These results are roughly the same as the results of previous European studies but differ from Hungarian research. In Bulgaria and Germany, the observed values for 9 and 8 markers were significantly lower than expected (Nikolov et al., 2009). In contrast, in Hungary one of the 9 markers in the previously studied cases differed significantly from HWE, and in the other study in 8 out of the 14 markers the observed heterozygosity value exceeded the expected one (Vernesi et al., 2003; Costa et al., 2012).

Determining the size and distribution of populations/subpopulations, examining the factors that separate them and the degree of genetic differences between them

After calculating the genetic indices of the population, they were grouped using Geneland software. According to the results, the most likely distribution is 2 populations, with a probability of nearly 80%. One group is located in the western part of the sampled area,

including Hungary except for its north-eastern region, and the other group includes the north-eastern part of Hungary and the foreign samples. The two groups are separated by a strong border, and in the south there is an "intrusion" where individuals that are typical in the northern group are found. For easier understanding, the samples were placed on a map provided by Google using ArcGis software (Figure 2).

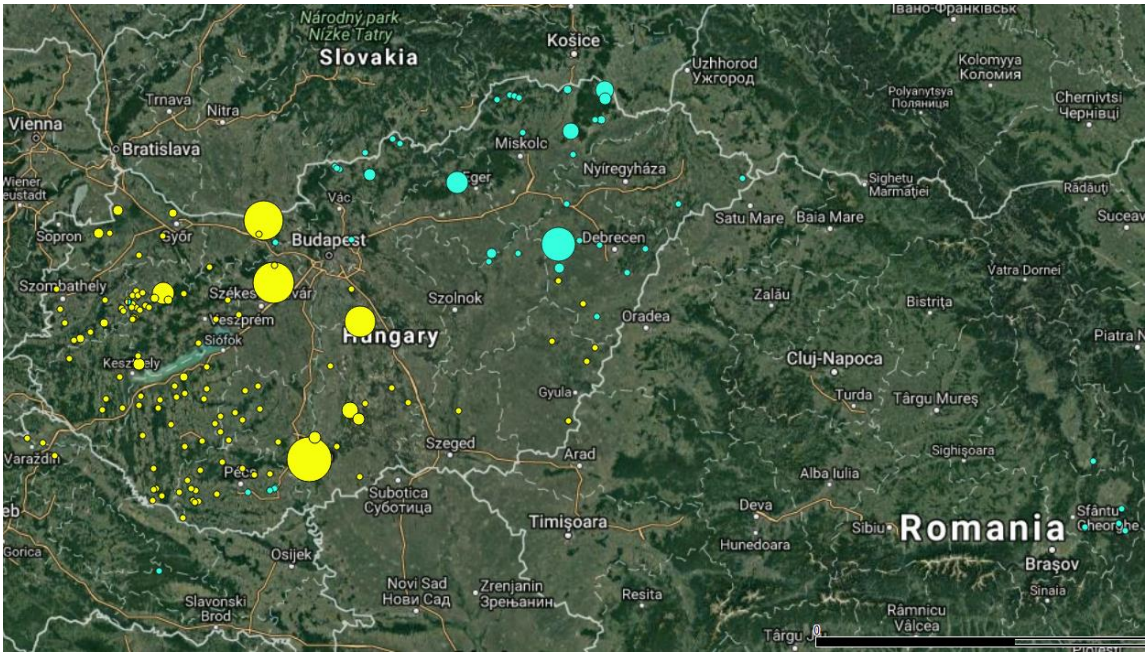

Figure 2: **Samples classified into subpopulations in ArcGis software**

Legend: blue circle: first group (n=147), yellow circle: second group (n=339). The size of the circles shows the number of samples collected in a single game management unit

The boundary line between the two subpopulations runs through Komárom-Esztergom, Pest, Jász-Nagykun-Szolnok and Hajdú-Bihar counties. There is only one individual that didn't fit its neighboring group, in Veszprém county. Based on our results the samples obtained from the neighboring countries fit well into the two subpopulations. This result is not surprising at all, because the political borders do not, or only partially coincide with the natural borders (for example the Danube in the western part of the Hungarian-Slovakian border, or the Drava on the Hungarian-Croatian border); however, these natural barriers do not constitute an impenetrable barrier for the wild boar. The southern border fence of Hungary was built on the Hungarian-Croatian border during the sampling period, so its effect has not yet shown. In the case of Romanian samples, the large geographical distance could have caused segregation, as the individuals sampled were collected more than 500km away from Hungary. The reason for the lack of segregation is presumably

the fact that wild boars also live between the two sampled groups, in locations from which DNA could not be collected.

After grouping the sample set it was necessary to determine how they are related to each other and what the difference between them is. The relationship between the groups is shown by the F_{st} value, therefore I calculated the value of this indicator, also using the Geneland software. In our case the F_{st} value is 0.03, which does not reach the level of moderate segregation (Fst>0.05), therefore the two groups are identified as two subpopulations of a single population. The relative genetic position of the subpopulations by principal component analysis is shown in the figure below (Figure 3):

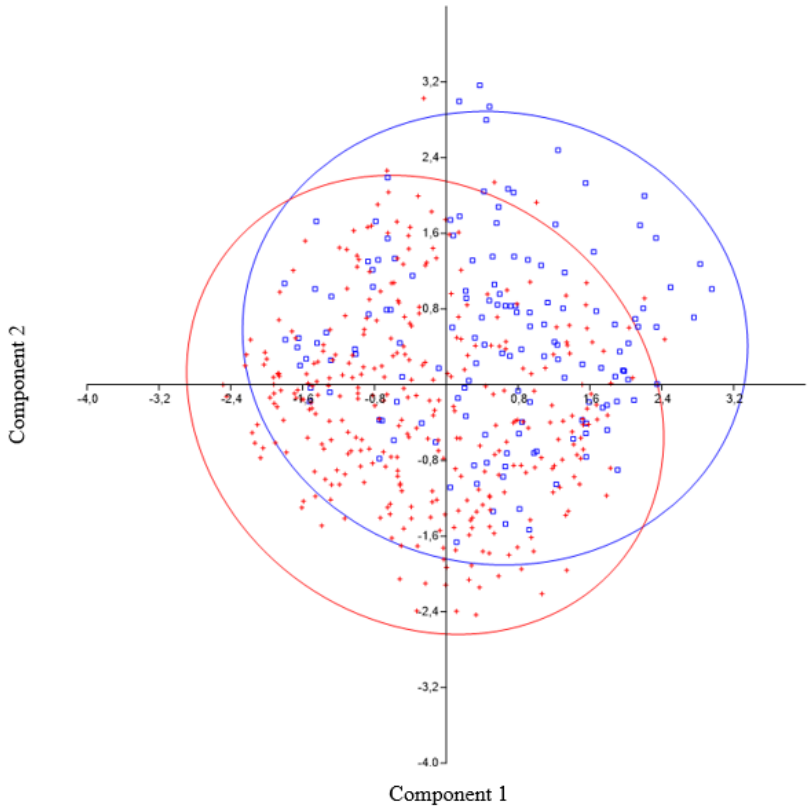

Figure 3: **Genetic location of the two subpopulations relative to each other by principal component analysis**

Legend: blue squares: individuals in the first group (n=147), red plus signs: individuals in the second group (n=339), colored circles are the 95% probability sets drawn around the groups

It can be clearly seen in the Figure that the two groups are not separated from each other, the vast majority of individuals are at the intersection of the two probability sets. The sizes of the sets do not differ significantly from each other, which means that the genetic diversity is nearly the same. This result confirms the results obtained by Geneland and GenAlEx, i.e. that we cannot speak of separate populations; however, there are differences between the two groups, which means subpopulations.

The simplest explanation for segregation between groups is geographical barriers. As a first step, we used the map presented earlier to look for different obstacles in the area between subpopulations, such as highways, lakes, rivers. As we could not find a clear geographic barrier, we used the Barrier software to search for separation lines. The results obtained by the program do not follow either the line of genetic separation we obtained or the real geographical conditions, therefore we looked for other possible causes. As the resettlement in the Carpathian Basin after the last ice age has come from multiple places as described for other species, we looked for traces of past effects. Finally, a hydrogeographic map from the early 1900s and a comparison of our results provided a possible explanation for the distribution. Figure 4 below shows that the north-eastern part of the floodplain, bordering the Mátra mountains roughly coincides with the location of the subpopulation marked in blue and the barrier marked „f" by the Barrier software. However, finding the exact reasons for segregation is also made difficult by the fact that there are more than 100 game reserves nationwide that can relocate individuals from elsewhere after licensing. Thus, traces of a previous relocation can significantly affect the results.

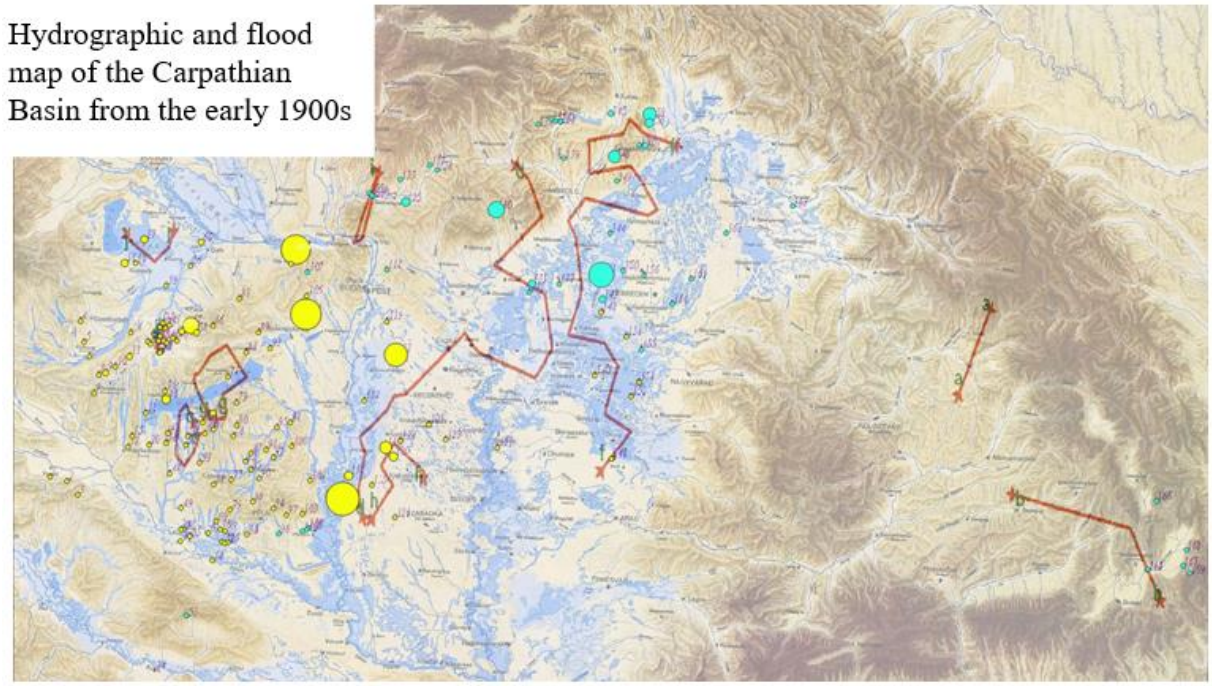

Figure 4: **Subpopulations and Barrier software results depicted on a flood map from the early 1900s**

Testing a marker set developed for individual identification on a large number of samples at the regional level

According to the statistical formula, the marker set we used can separate a total of $9.65*10^{18}$ (approx. 10 trillion) individuals for the allele numbers obtained, but the genetic profiles of two of the 486 individuals we examined are completely identical on every allele. Three possible explanations were found for the reason of the similarity: in both individuals, 2 markers (PigSTR11B and PigSTR1B) did not give results even after repetition, therefore it cannot be excluded that there is a difference in those parts. The site where the samples were collected was Borota, thus they may be identical twins, which may also occur in wild boars (Náhlik et al., 2013). In this case, the individuals actually have the same genetic profile. Finally, although the evaluation was also re-checked, it is possible that the two individuals match due to an evaluation error.

Population dynamics studies of wild boar populations in the Carpathian Basin

Investigation of the bottleneck effect and previous genetic shrinkage of the groups

Data from the microsatellite marker set were used to test the existence of the bottleneck effect. During analyses two methods were used: the stepwise mutation model (SMM) and the two-phase mutation model (TPM). Both statistical methods show that all 13 markers are in mutation-drift equilibrium, and the mode-shift test also shows a normal (L-shaped) distribution, which means that no genetic regression has has taken place recently. However, post-hoc tests showed significant differences from HWE, with less than expected heterozygous individuals in 13 and 11 loci by the SMM and TPM models, respectively, which was also confirmed by the Wilcoxon rank sum test (SMM=0.0006; TPM=0.00201). Based on the results it can be stated that in the distant past the Hungarian wild boar population was affected by genetic shrinkage.

Kinship analyses of the wild boar sample set tested

The indicators of the sample set were examined for the following relationships: siblings, half-siblings, maternity and paternity. Analyzes were performed using the microsatellite data (n=486) and the combined data (n=422). The sibling study revealed the presence of 23 and 22 sibling pairs respectively, with a probability of more than 90%. Between the two results 7 sibling pairs overlap at this probability level. In addition, according to the results of the combined marker set, two sibling groups of 3 individuals each were sampled. We then checked the geographical distances from each other, which gave a surprising result: although in theory the siblings live next to each other, at most a few 10 km apart, our results show otherwise. In this comparison, the results of the combined marker set seem more reliable, as the average distance is only 61.3 km compared to the 110 km result of the STR set. In both cases the standard deviation is above the average, caused by some very distant siblings. These outliers can be caused by the relocation of wild boars, which is a common activity between wild boar hunting gardens, and despite the high probability value, a false result due to a genotyping error cannot be ruled out either.

The half-sibling study revealed the presence of 18 and 29 half-sibling pairs, respectively, with a probability of more than 90%. Between the two results 6 pairs overlap at this probability level. According to the results of the microsatellite marker set, three half-sister groups of more than 2 individuals were sampled, and the combined marker set showed 5 groups, of which the largest one included 5 individuals. Checking the geographical

distance gave a result different from that before: in this case the average distance of the microsatellite set is 133 km compared to the result of 192 km of the combined marker set. However, when the samples are checked on the map, it is clear that the distant individuals come from 3 core areas, which are approximately 120, 240 and 320 km apart from each other. In addition, two of the three core areas are large-scale game management units, including wild boar hunting gardens, therefore in this case the relocations explain the large geographical distance between the half-siblings. The parental study showed only mother-offspring relationships when the combined set was used, with 2 mothers and 5 and 4 offsprings. The offsprings are the same as the two large groups in the half-sibling study, consequently the common parent of these groups was the sow. Therefore the large distances from each other can presumably be traced back to relocation.

Testing of a set of markers developed for tracing of wild boar meat products on natural groups; hybridization studies

Investigation of the accuracy of the developed marker set by bioinformatics

The success of marker development is well exemplified by the following comparison, in which we compared the results of our own marker set with the latest and most detailed research available by Lorenzini et al. (2020). The W3 marker was evaluated in two ways, as the marker was designed for the reference genome SSC10.2, but since then a new genome (SSC11.1) has become available, where the region used has been changed, therefore the results of the marker seem fragmented in the software. In one evaluation the most frequently obtained fragment color, and in the other the average value of the colors was recorded (Table 1).

Table 1:

Comparison of the accuracy of different markers developed for hybridization

Breed	W1	W2	W3 (legtöbb)	W3 (átlag)	MC1R2/3	MC1R1	NR6A1
Angler schatterswein	100	-	-	-	-	-	100
Angler schatterswein	100	100	100	50	100	100	100
Bunte bentheimer	100	-	-	0	-	-	100
Berkshire	100	0	50	50	100	0	50
Berkshire	100	100	50	50	100	0	100
British saddleback	100	100	50	50	50	100	100
British saddleback	100	100	0	50	50	100	-
Wild boar	50	0	100	50	0	0	100
Wild boar	100	100	100	50	0	0	100
Wild boar	100	100	100	50	-	-	-
Wild boar	100	0	100	50	0	-	100
Wild boar	50	50	100	50	0	0	100
Wild boar	100	0	0	50	0	0	100
Wild boar	100	0	100	50	0	-	-
Wild boar	100	100	100	50	0	-	-
Wild boar	100	100	100	50	0	0	0
Wild boar	100	50	100	50	0	0	100
Wild boar	100	-	100	50	0	-	0
Wild boar	100	100	100	50	0	0	50
berkshire	100	100	50	50	100	0	-
Berkshire	100	100	100	50	100	-	100
Berkshire	0	100	100	50	-	-	100
berkshire	50	100	50	50	100	0	50
Sum	2050	1400	1650	1050	700	300	1450
Number	23	20	21	22	19	14	18
Accuracy (%)	89.13	70	78.57	47.73	36.84	21.43	80.56

The results show that the accuracy of MC1R polymorphisms is extremely poor, only 21.43% and 36.85%. This is followed by the accuracy of the W3 and W2 InDel markers, and the two best results are given by the NR6A1 and W1 markers.

Examination of the hybridization level of the Hungarian wild boar population

Hybridization surveys were performed on 422 wild boar samples and a reference database of 120 domestic pigs (1: pietrain, 2: hampshire, 3: large white, 4: H39 x large white, 5: landrace, 6: duroc, 7: duroc x mangalica, 8: blonde mangalica, 9: swallow-bellied mangalica, 10: red mangalica, n=12 in each group), and 10 known wild boar-domestic pig hybrids were also included. The Bayesian clustering results given by the Structure software using the 3 InDel markers are shown in the figure below (Figure 5):

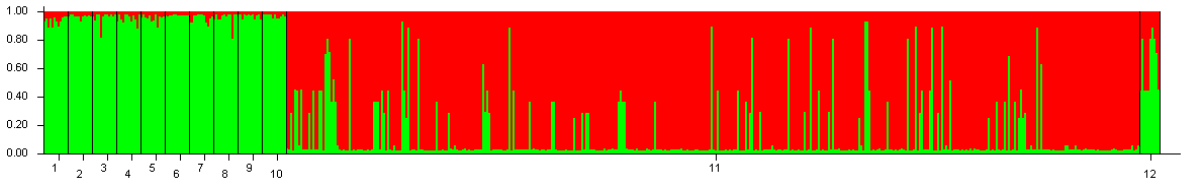

Figure 5: **Hybridization characteristics based on 3 InDel markers**

Legend: groups 1-10: domestic pigs, group 11: wild boar, group 12: known hybrids

Each vertical line represents an individual

Color explanation: green: domestic pig, red: wild boar

The results clearly show that the software identified all 120 individuals in the first 10 groups (domestic pigs) as pure domestic pigs, 336 of the wild boar samples were evaluated as wild boar, 71 as hybrids and 18 as domestic pigs, and 6 of the known hybrids were classified as hybrids and 4 as domestic pigs.

In the next step, the marker sets were combined, and this data set was also clustered in the Structure software (Figure 6):

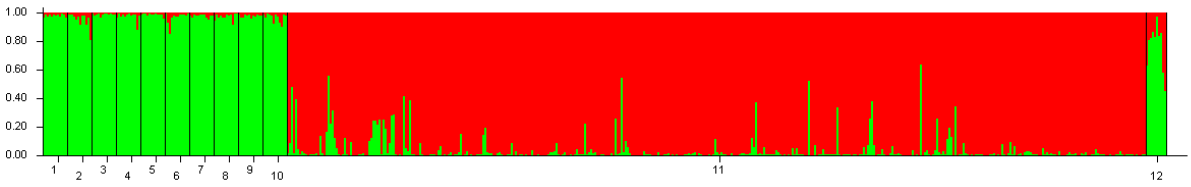

Figure 6: **Hybridization characteristics based on combined marker sets**

Legend: groups 1-10: domestic pigs, group 11: wild boar, group 12: known hybrids

Each vertical line represents an individual

Color explanation: green: domestic pig, red: wild boar

Based on the results of the combined sets, there was no change in the first 10 groups compared to the previous ones, as all 120 domestic pigs were identified as domestic pigs;

20 wild boars were classified as hybrids, and from the known hybrids 3 were evaluated as hybrids and 7 as domestic pigs.

Testing the reliability of the hybridization marker set

After comparing the results, it can be stated that the marker set we made is "stricter" than the validating (STR) set, as it also identifies wild boars as hybrids and domestic pigs, and also indicates some of the known hybrids as domestic pigs. The set does not classify domestic pigs as wild boars in any of the cases, but three individuals were classified as hybrids. Overall, the hybridization marker set gave the true result in 472 cases, which is 85.5% of the total sample set. It was „too strict" (the error that could be corrected by re-examination using the validation marker set) in 13.41%, and „too permissive" (the error that couldn't be corrected) in 1.09%. It can be stated that the marker set gives a false positive result in approximately 14% of the cases, which can be eliminated if the individuals in question are also examined with the STR marker set. We obtained a false negative result of only 1%, which can be said to be a particularly good result because the InDel marker set was originally developed as a fast and inexpensive pre-screening method. All in all, the results show that the marker development was successful.

Testing the results of the hybridization marker set by adding simulated genotypes

As a final step in the hybridization studies, wild boars with the purest genotype and domestic pigs were crossed for 2 generations using the Hybridlab software. The Bayesian clustering of the sample set extended with virtual individuals was also performed in two ways, with InDel set only, and with combined marker set.

By the results of the InDel set the software identified all domestic pigs as domestic pigs. Of the wild boar samples 339 were identified as wild boars, that is 3 individuals more than before. 68 were evaluated as hybrids and 18 as domestic pigs. From the known hybrids only 3 were recognized as domestic pigs in this case, and 7 were identified as hybrids. Based on the combined results, all domestic pigs were also classified in their own cluster. Of the wild boars 27 animals were classified as hybrids, which is 7 individuals more than before. No wild boar was identified as a domestic pig. The results of the known hybrids were unchanged: in 7 cases they were classified as domestic pigs and in 3 cases as hybrids. The summarized results for the virtual genotypes are shown in Table 2 below:

Table 2:

Classification results of simulated individuals by two methods according to hybridization rate

Legend: DP: domestic pig, WB: wild boar

Color explanation: green: identical genotype, yellow: correctable error, red, uncorrectable error

		W			STRW		
N	**Rate of hybridization (DP/WB%)**	**DP**	**hybrid**	**WB**	**DP**	**hybrid**	**WB**
24	75-25%	21	3	0	20	4	0
36	50-50%	15	20	1	13	21	2
24	25-75%	6	12	6	2	12	10

The results show that the number of individuals is shifted towards domestic pigs in all cases compared to the expected distribution. From this it can be concluded that the percentage at which the individual is classified as a hybrid is too strict in our case, as even a few domestic alleles are evaluated with too much emphasis by the software.

In summary, based on our results, 2.84% of the Hungarian wild boar population is hybridized. Compared to previous studies with other STR marker sets, this number is average. Two previous types of research did not detect hybrid individuals (Nikolov et al., 2017; Sprem et al., 2014). In a study on Italian wild boars using 20 markers on 164 individuals, a rate of 2.44% was found (Lorenzini et al., 2020). In a multi-country study, Frantz et al. (2013) studied 697 boars with a set of 14 markers, and found that 6.32% of wild boars showed signs of hybridization. Research based on SNP mutations affecting fur colour (MC1R and NR6A1) has shown a much higher rate of hybridization; however, these studies look at much fewer DNA segments, thus the results are less accurate than studies using numerous STR markers and multiple polymorphisms.

5. New scientific results

1. We were the first in Hungary to perform a genetic study with such a large scale of wild boars using STR markers. We determined the degree of genetic difference (7.62 allele/marker) and the hybridization rate (2.84%) of the Hungarian wild boar population. Compared to other European studies, both values are within the previously established range.
2. We found that the wild boar population of Hungary is divided into two subpopulations (north-east n=147, rest of the country n=339). The boundary line is presumably the remnant of the last ice age and water management, as there is no significant barrier between the two subpopulations.
3. We have proved that with the STR marker set we used in Hungary, individual identification can be performed on wild boars with a minimum of 99,79% accuracy.
4. By genetic studies we have shown the presence of a bottleneck effect (significant heterozygote deficiency in 13/13 markers using SMM method,and in 13/11 markers with TPM method which was also confirmed by the Wilcoxon rank sum test (SMM=0.0006; TPM=0.00201)), and the absence of current genetic shrinkage events in the Hungarian wild boar population.
5. We tested and proved the reliability of the previously developed InDel hybridization marker set in the wild boar population, and also determined its necessity in addition to the widely used methods (MC1R and NR6A1). The accuracy of 85.5%, together with a false negative result of only 1% gives game managers the means to perform highly reliable genetic testing on wild boars much faster and at a lower cost than before.

6. Practical use of the results

1. With a comprehensive sampling of wild boars in Hungary and the determination of their genetic diversity and structure, we have created a database that provides a useful basis for future research, or comparison for population genetics, diversity or hybridization studies.
2. Knowing the size and location of subpopulations, it is possible to develop comprehensive, large-scale game management plans for both hunting and control of ASF.
3. The use of the individual identification STR marker set provides a solution for dealing with various criminal cases such as poaching or trophy counterfeiting using modern genetic methods.
4. Using the InDel (and the STR, if necessary) marker set, it is possible to examine wild boar groups (in game gardens, game parks or farms) in order to filter out hybrid individuals. Thus, the groups can stay genetically pure wild boars, which can also affect the phenotype and meat quality. In all cases, we recommend sampling wild boars relocated into game hunting gardens, which can be tested at the time of sale, thus certifying the genetic background and quality of the individual sold.

7. Bibliography

CORNUET, J.M. & LUIKART, G. 1996: Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144, 2001-2014. pp.

COSTA, V., PÉREZ-GONZÁLEZ, J., SANTOS, P., FERNÁNDEZ-LLARIO, P., CARRANZA, J., ZSOLNAI A., ANTON I., BUZGÓ J., VARGA GY., MONTEIRO, N., BEJA-PEREIRA, A. 2012: Microsatellite markers for identification and parentage analysis in the European wild boar (*Sus scrofa*). BMC Research Notes 5(479): 6 pp.

FEGYVERNEKI S. 2011: Valószínűség-számítás és matematikai statisztika. Miskolc, 150 pp.

GUILLOT, G., RENAUD, S., LEDEVIN, R., MICHAUX, J., CLAUDE, J. 2012: A unifying model for the analysis of phenotypic, genetic and geographic data. Systematic Biology 61(5): 897-911. pp.

HAMMER, Ø., HARPER, D.A.T., RYAN, P.D. 2001: PAST: paleotological statistics software package for education and data analysis. Palaeontologica Electronica 4(1). 9 pp.

JONES, O. & WANG, J. 2010: COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10: 551-555. pp.

LIN, Y-C., HSIEH, H-M.,LEE, J-C-I., HSIAO, C-T., LIN, D-Y., LINACRE, A., TSAI, L-C. 2014: Establishing a DNA identification system for pigs (Sus scrofa) using a multiplex STR amplification. Forensic Science International: Genetics 9, 12-19. pp.

LORENZINI, R., FANELLI, R., TANCREDI, F., SICLARI, A., GAROFALO, L. 2020: Matching STR and SNP Matching STR and SNP genotyping to discriminate between wild boar, domestic pig, and their recent hybrids for forensic purposes. Scientific Reports, Vol 10(3188).

MANNI, F., GUÉRARD, E., HEYER, E. 2004: Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by „Monmonier's algorithm". Human Biology, 76(2), 173-190. pp.

NÁHLIK A., SÁNDOR GY., TARI T. 2013: A vaddisznó (Sus scrofa) szaporulatának alakulása egy szabadterületi populációban. Erdészettudományi Közlemények 3(1): 261-269. pp.

NIELSEN, E.E.G., BACH, L., KOTLICKI, P. 2006: Hybridlab (version 1.0): a program for generating simulated hybrids from population samples. Molecular Ecology Notes 6, 971-973. pp.

NIKOLOV, I.S., STOECKLE, B.C., MARKOV, G., KUEHN, R. 2017: Substantial hybridization between wild boars (Sus scrofa scrofa) and East Balkan pigs (Sus scrofa f. domestica) in natural environment as a result of semi-wild rearing in Bulgaria. Czech Journal of Animal Science, Vol 62(1), 1-8. pp.

PEAKALL, R. & SMOUSE, P.E. 2012: GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioiformatics 28(19), 3 pp.

PRITCHARD, J.K., STEPHENS, M., DONNELLY, P. 2000: Inference of population structure using multilocus genotype data.Genetics155: 945–959. pp.

ROBINSON, J.T., THORVALDSDOTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E.S., GETZ, G., MESIROV, J.P. 2011: Integrative Genomic Viewer. Nature Biotechnology 29, 24-26.

SPREM, N., SALAJPAL, K., SAFNER, T., DIKIC, D., JURIC, J., CURIK, I. DIKIC, M., CUBRIC-CURIK, V. 2014: Genetic analysis of hybridization between domesticated endangered pig breeds and wild boar. Livestock Science, Vol 162, 1-4. pp.

SZEMETHY D., MIHALIK B., FRANK K., NAGY T., ÚJVÁRY D., KUSZA SZ., SZEMETHY L., BARTA E., STÉGER V. 2020: Development of wild boar species-specific DNA markers for a potential quality control and traceability method in meat products. Food Analytical Methods 2020/14, 18-27. pp.

VERNESI, C., CRESTANELLO, B., PECCHIOLI, E., TARTARI, D., CARAMELLI, D., HAUFFE, H., BERTORELLE, G. 2003: The genetic impact of demographic decline and reintroduction in the wild boar (Sus scofa): A microsatellite analysis. Molecular Ecology 12, 11 pp.